

A DSMT Based Combination Scheme for Multi-Class Classification

Nassim Abbas, Youcef Chibani, Zineb Belhadi and Mehdi Hédir

Speech Communication and Signal Processing Laboratory,
Faculty of Electronics and Computer Science

University of Science and Technology Houari Boumediene (USTHB),

32, El Alia, Bab Ezzouar, 16111, Algiers, Algeria

nabbas@usthb.dz, ychibani@usthb.dz, zbelhadi@usthb.dz, mhedir@usthb.dz

Abstract—This paper presents a new combination scheme for reducing the number of focal elements to manipulate in order to reduce the complexity of the combination process in the multi-class framework. The basic idea consists in using of p sources of information involved in the global scheme providing p kinds of complementary information to feed each set of p one class support vector machine classifiers independently of each other, which are designed for detecting the outliers of the same target class, then, the outputs issued from this set of classifiers are combined through the plausible and paradoxical reasoning theory for each target class. The main objective of this approach is to render calibrated outputs even when less complementary responses are encountered. An inspired version of Appriou's model for estimating the generalized basic belief assignments is presented in this paper. The proposed methodology allows decomposing a n -class problem into a series of n -combination, while providing n -calibrated outputs into the multi-class framework. The effectiveness of the proposed combination scheme with proportional conflict redistribution algorithm is validated on digit recognition application and is compared with existing statistical, learning, and evidence theory based combination algorithms.

Keywords—Conflict management; Dezert-Smarandache theory; Handwriting digit recognition; Support Vector Machines.

I. INTRODUCTION

Nowadays a large number of classifiers and methods of generating features is developed in various application areas of pattern recognition [1], [2]. Nevertheless, it failed to underline the incontestable superiority of a method over another in both steps of generating features and classification. Rather than trying to optimize a single classifier by choosing the best features for a given problem, researchers found more interesting to combine the recognition methods [2]. Indeed, the combination of classifiers allows exploiting the redundant and complementary nature of the responses issued from different classifiers.

Researchers have proposed an approaches for combining classifiers increasingly numerous and varied, which led the development of several schemes in order to treat data in different ways [2]. Generally, three approaches for combining classifiers can be considered: parallel approach, sequential approach and hybrid approach [2]. Furthermore, these ones can

be performed at a class level, at a rank level, or at a measure level [3].

However, with the existence of the constraints corresponding to the joint use of classifiers and methods of generating features, an appropriate operating method using mathematical approaches is needed, which takes into account two notions: uncertainty and imprecision of the responses of classifiers. In general, the most theoretical advances which have been devoted to the theory of probabilities are able to represent the uncertain knowledge but are unable to model easily the information which is imprecise, incomplete, or not totally reliable. Moreover, they often lead to confuse both concepts of uncertainty and imprecision with the probability measure. Therefore, a new original theories dealing with uncertainty and imprecise information have been introduced, such as the fuzzy set theory [4], evidence theory [5], possibility theory [6] and, very recently, the theory of plausible and paradoxical reasoning in [7], [8], [9]. The Dezert-Smarandache theory (DSMT) of plausible and paradoxical reasoning was elaborate by Jean Dezert and Florentin Smarandache for dealing with imprecise, uncertain and paradoxical sources of information. Thus, the main objective of the DSMT was to introduce combination rules that would allow to correctly combining evidences issued from different information sources, even in presence of conflicts between sources or in presence of constraints corresponding to an appropriate model (i.e. free or hybrid DSMT models [7]). The DSMT is justified in many kinds of applications [7], [8], [9].

Indeed, the DSMT has a feasible computational complexity for industrial uses which are considered as problems of small dimension [10], [11]. In contrast, the extension of this theory into the multi-class framework has the problem of their applicability in view of the high computational complexity. This is closely related with the number of elements to be processed in the framework of this theory, which follows the sequence of Dedekind's numbers [12]. Try to use the free DSMT model, considering the set of all subsets of the original classes (but under the union and the intersection operators), is not easy and becomes untractable for more than 6 elements in the discernment space [13].

In this paper, we propose an effective combination scheme of one-class classifiers in a general belief function framework

by incorporating an intelligent learning technique for reducing the number of focal elements. This allows us to reduce drastically the computational complexity of the combination process and to extend specially the applicability of DSMT into the multi-class classification framework. Indeed, the objective of this work is neither to choose the kind of the one-class classifier, but only to illustrate from a practical application the advantage of this new combination scheme for the real-time implementation purpose.

In Section II, this paper will deal with the related works. Theoretic formulation of Proportional Conflict Redistribution (PCR6) combination rule and the way it can be extended for solving the multi-class classification problem are presented in Section III. We give in Section IV a multi-class classification scheme based on belief function theories. The database of the isolated handwritten digits, methods used for generating features and algorithm used for OC-SVM models validation are described in Section V. The experimental and statistical results are summarized in Section VI.

II. RELATED WORKS

Dezert and Smarandache [14] proposed a first work for ordering all elements generated using the free DSMT model for matrix calculus such as [15], [16] made in DST framework. But, this proposition has limitations since in practical applications it is better to only manipulate the focal elements [17], [18], [19], [3].

Hence, few works have been focused already on the computational complexity of the combination algorithms formulated in DSMT framework. Djiknavorian and Grenier [18] showed that there's a way to avoid the high level of complexity of DSMT hybrid (DSMH) combination algorithm by designing a such code that can perform a complete DSMH combination in very short period of time. However, even if they have obtained an optimal process of evaluating DSMH algorithm, first some parts of their code are really not optimized and second it has been developed only for a dynamic fusion. Martin [20] further proposed a practical codification of the focal elements which gives only one integer number to each part of the Venn diagram representing the discernment space. Contrary to the Smarandache's codification [13] used in [21] and the proposed codes in [18], author thinks that the constraints given by the application must be integrated directly in the codification of the focal elements for getting a reduced discernment space. Therefore, this codification can drastically reduce the number of possible focal elements and so the complexity of the DST as well as the DSMT frameworks. A disadvantage of this codification is that the complexity increases drastically with the number of combined sources especially when dealing with a problem in the multi-class framework. To address this issue, Li *et al.* [22] proposed a criterion called evidence supporting measure of similarity (ESMS), which consists in selecting, among all sources available, only a subset of sources of evidence in order to reduce the complexity of the combination process, but this criterion has been justified for only a two-class problem. However, the complexity of reducing both the number of combined sources and the size of the discernment space are research challenges that still need to be addressed.

III. THEORETIC FORMULATION OF PCR6 COMBINATION RULE AND APPLICATION TO MULTI-CLASS CLASSIFICATION

We denote by p sources S_1, S_2, \dots, S_p of information corresponding to measures issued from p one-class classifiers, which are involved to deal with a one-class problem. For each source, two classes θ_i and $\bar{\theta}_i$ are considered being the *target* class and their *complementary* class, respectively.

To generate the combined masses from all the p one-class classifiers, a set of elements is defined namely $G^{(i)} = \{A_1, \dots, A_{Car(G^{(i)})}\}$, which attributes for each element A_j ($j = 1, \dots, Car(G^{(i)})$) a "belief" value $m_k(A_j)$ associated to each source S_k taking the value in the range $[0, 1]$ and verifying $\sum_{j=1}^{Car(G^{(i)})} m_k(A_j) = 1$, such that $Car(G^{(i)})$ is the cardinal of the set $G^{(i)}$ and $k \in \{1, 2, \dots, p\}$ defines the index of the corresponding source.

According the finite set of hypotheses, a combination rule is then performed on elementary masses for generating partial masses from all the p sources as follows:

$$m_c^{(i)}(A) \equiv [m_1 \oplus \dots \oplus m_p](A), A \in G^{(i)}, \quad (1)$$

where $m_c^{(i)}(A)$ is the partial mass associated to the elementary or compound hypothesis A of the combined sources and verifying $\sum_{j=1}^{Car(G^{(i)})} m_c^{(i)}(A_j) = 1$.

Hence, the choice of an appropriate combination rule depends on the set of predefined hypotheses. Example of such approaches is Proportional Conflict Redistribution (PCR6) rule based on DSMT.

A. Combination Rule Based On the DSMT

Let $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ the *discernment space* of the multi-class classification problem under consideration having n *exhaustive elementary* hypotheses θ_i , which are not necessarily mutually *exclusive* in DSMT. The main concept of the DSMT is to distribute basic belief assignment of certainty over all the composite propositions built from elements of subset of Θ , namely $\Theta^{(i)} = \{\theta_i, \bar{\theta}_i\}$, with \cup (Union) and \cap (Intersection) operators instead of making this distribution over the elementary or union hypothesis only. Therefore, the hyperpowerset for two hypotheses (classes) belonging to $\Theta^{(i)}$ is defined as $G^{(i)} = D^{\Theta^{(i)}} = \{\emptyset, \theta_i, \bar{\theta}_i, \theta_i \cup \bar{\theta}_i, \theta_i \cap \bar{\theta}_i\}$. The DSMT uses generalized basic belief mass, also known as the generalized basic belief assignment (gbba) computed on hyperpowerset of $\Theta^{(i)}$, which is defined as: $m_k(\emptyset) = 0$ and $m_k(\theta_i) + m_k(\bar{\theta}_i) + m_k(\theta_i \cup \bar{\theta}_i) + m_k(\theta_i \cap \bar{\theta}_i) = 1$, such that $m_k(\theta_i \cup \bar{\theta}_i)$ defines the mass of the ignorance, $m_k(\theta_i \cap \bar{\theta}_i)$ represents the mass of the conflict (or paradoxical information), and $k \in \{1, 2, \dots, p\}$. The way the conflicting mass is

redistributed yields to several versions of a Proportional Conflict Redistribution (PCR) rules [23], [19]. From PCR1 to PCR2, PCR3, PCR4, PCR5 one increases the complexity of the rules and also the exactitude of the redistribution of conflicting masses. The combination rule (PCR5) proposed by [23] for two sources is mathematically one of the best for the proportional redistribution of the conflict applicable in the context of the DSMT. Martin and Osswald have proposed the following alternative rule to PCR5 for combining more than two sources altogether (i.e. p upper then 3). This new rule denoted PCR6 does not follow back on the track of conjunctive rule as PCR5 general formula does, but it gets better intuitive results. For $p = 2$ PCR5 and PCR6 coincide. The combined partial gbba $m_{PCR6}^{(i)}$ obtained from p generalized belief assignments $m_1(\cdot), \dots, m_p(\cdot)$ by means of the PCR6 rule [19] is defined as:

$$m_c^{(i)}(A) = m_{PCR6}^{(i)}(A) = \begin{cases} 0 & \text{if } A \in \Phi, \\ m_{\wedge}^{(i)}(A) + \sum_{k=1}^p m_k(A)^2 \sum_{\substack{j=1 \\ (Y_{\sigma_k(1)}, \dots, Y_{\sigma_k(p-1)}) \in (G^{(i)})^{p-1}}}^{p-1} L_k & \text{otherwise.} \end{cases} \quad (2)$$

Where

$$L_k = \frac{\prod_{j=1}^{p-1} m_{\sigma_k(j)}(Y_{\sigma_k(j)})}{m_k(A) + \sum_{j=1}^{p-1} m_{\sigma_k(j)}(Y_{\sigma_k(j)})}, \quad (3)$$

$\Phi = \{\Phi_M, \emptyset\}$ is the set of all relatively and absolutely empty elements, Φ_M is the set of all elements of $G^{(i)}$ which have been forced to be empty in the hybrid model M defined by the exhaustive and exclusive constraints, \emptyset is the empty set,

the denominator $m_k(A) + \sum_{j=1}^{p-1} m_{\sigma_k(j)}(Y_{\sigma_k(j)})$ is different to zero,

and where σ_k counts from 1 to p avoiding k , i.e.:

$$\sigma_k(j) = \begin{cases} j & \text{if } j < k, \\ j+1 & \text{if } j \geq k. \end{cases} \quad (4)$$

Here, $m_{\wedge}^{(i)}(A)$ corresponds to the classical DSMT rule on the free DSMT model [24], which is defined as:

$$m_{\wedge}^{(i)}(A) = \begin{cases} 0 & \text{if } A = \emptyset, \\ \sum_{\substack{B_1, B_2, \dots, B_p \in G^{(i)} \\ B_1 \cap B_2 \cap \dots \cap B_p = A}} \prod_{k=1}^p m_k(B_k) & \text{otherwise.} \end{cases} \quad (5)$$

B. Effective Combination Scheme of One-Class Classifiers

For the computation of the global combined mass $m_c(\cdot)$, the direct use of PCR6 combination rule on D^{Θ} yields the computation cost that increases drastically with n and even may be computationally prohibitive, especially when we have a huge number of elements belonging to D^{Θ} when dealing with gbba's within the DSMT framework (i.e. $Card(D^{\Theta}) = d(n)$, such that $d(n)$ is the Dedekind's number of n). However, for our multi-class problem, because of a special feature that the separation of data according to One Against All (OAA) approach [25] of such classifier, we can render the data highly unbalanced for each two-class problem. Hence, the need to use an one-class classifier which is able to distinguish the samples of the target class θ_i from other outliers belonging to its complementary class $\bar{\theta}_i$, $i = 1, \dots, n$. From this principle, we propose a combination scheme which allows decomposing a n -class problem into a series of n -combination, whose reasoning for each combination is performed from the subset $\Theta^{(i)}$, $i = 1, \dots, n$, instead of the reference space Θ . Hence, we propose a combination scheme which uses a complementary features captured by the different sources of information S_k , $k = 1, \dots, p$, from the input probe data to feed the one-class classifiers $OC-C_{ik}$, $k = 1, \dots, p$, that operate independently of each other for each target class θ_i , $i = 1, \dots, n$, and then the partial opinions (i.e. transformed measures) provided from these classifiers will be combined all together through an appropriate rule in the subset $G^{(i)}$, $i = 1, \dots, n$. Finally, all the n partial combined masses will be incorporated in a unique module for the task of decision making. Table I gives a comparison of $Card(2^{\Theta})$ and $Card(D^{\Theta})$ with respect to

$Card(F) = \sum_{i=1}^n Card(G^{(i)})$, which defines the cardinal of set of

all focal elements $F = \{G^{(i)}, i = 1, \dots, n\}$ obtained according to our computing method within DSMT framework, as follows:

TABLE I
CARDINALITY OF COMBINATION SPACE

$Card(\Theta) = n$	$Card(2^{\Theta})$	$Card(D^{\Theta})$	$Card(F)$
2	4	5	5
3	8	19	11
4	16	167	14
5	32	7580	17
6	64	7828353	20

In this way, we can reduce drastically, within DSMT framework, the number of focal elements from $Card(D^{\Theta}) = d(n)$ to $Card(F) = 3 \times n + 2, \forall n \geq 3$.

IV. MULTI-CLASS CLASSIFICATION SCHEME BASED ON BELIEF FUNCTION THEORIES

The proposed multi-class classification scheme incorporates mainly four modules: i) one-class support vector machine (OC-SVM) classification, ii) transformation of the normalized OC-SVM outputs into belief assignments using estimation technique based on a modified version of Appriou's model, iii) combination of masses through an algorithm based on belief function theories and iv) decision making.

A. Classification Based On OC-SVM

In the proposed combination scheme, we take the path of OC-SVMs, enabling us to incorporate an intelligent learning technique to efficiently avoid both the closed set and good distribution assumptions in the multi-class classification framework. In the following, we briefly review the concept learning with one-class SVM.

Review of OC-SVM: Schölkopf *et al.* [26] proposed OC-SVM classifier by modifying the standard support vector machines initially introduced by Vapnik [27]. The pattern classification approach using OC-SVM has been successfully used for many applications as biometric verification [28], [29]. This classifier is an unsupervised learning algorithm, which only requires the learning of the target class samples. In fact, the pattern classification through OC-SVM consists of defining a boundary around the target class, such that it accepts as many of the target samples as possible, while minimizing the chance of accepting outliers.

The concept of the OC-SVM seeks to find an hyper sphere in which the most of learning data are included into a minimum volume. More specifically, the objective of the OC-SVM is to estimate a function $f_{OC}(x)$ that encloses the most of learning data into a hyper sphere $R_x = \{x \in \mathbb{R}^d, f_{OC}(x) > 0\}$ with a minimum volume where d is the size of feature vector [26]. Hence, the decision function $f_{OC}(x)$ is given as [26]:

$$f_{OC}(x) = \sum_j^{Sv} \alpha_j K(x, x_j) - \rho. \quad (6)$$

where Sv is the number of support vectors x_j form the training dataset, α_j are Lagrange multipliers, such that

$$0 \leq \alpha_j \leq \frac{1}{vm}, \quad m \text{ is the cardinal of training dataset, } v \text{ is the}$$

percentage of data considered as outliers, ρ defines the distance of the hyper sphere from the origin, and $K(.,.)$ defines the OC-SVM kernel that allows projecting data from the original space to the feature space [30].

A pattern x is then accepted when $f_{OC}(x) > 0$. Otherwise, it is rejected. Various kernel functions can be used as polynomial, Radial Basis Function (RBF) or multilayer perceptron [27]. Generally, the RBF kernel is used for its better performance, which is defined as:

$$K(x, x_j) = \exp\left(-\gamma \|x - x_j\|^2\right). \quad (7)$$

where γ is the kernel parameter.

In the following, we show how the OC-SVM based concept learning can be extended to construct multi-class OC-SVM with multiple hyper spheres.

Extension of OC-SVM for Constructing Multi-Class OC-SVM: Basically, the OC-SVM classifier have been conceived to deal with an one-class classification problem [26]. Their extension to multi-class scenario may provide uncalibrated outputs for some classifiers. In this paper, we use a sigmoid transformation for mapping the reassigned output, using logarithmic function, of different OC-SVM classifiers to probabilities as follows:

$$P_k(\theta_i/x) = \frac{1/Z_k}{1 + \exp(-g_{ik}(x))}, \quad (8)$$

where $Z_k = \sum_{i=0}^{n-1} [1 + \exp(-g_{ik}(x))]^{-1}$ represent normalization

factors that are introduced in the probabilistic framework in order to respect the normality condition (i.e. $\sum_{\theta_i \in \Theta} P_k(\theta_i/x) = 1$),

P_k are the posterior probabilities issued from the source of information $S_k, k=1, \dots, p$. Thus, the term $g_{ik}(x)$ represents the reassigned output of the i -th OC-SVM classifier, $i=1, \dots, n$, which is defined for a given pattern x as:

$$g_{ik}(x) = -\log \left[\sum_j^{Sv_{ik}} \alpha_j K(x, x_j) \right] + \log(\rho_{ik}), \quad (9)$$

Sv_{ik} and ρ_{ik} are the number of support vectors and the distance of the hyper sphere from the origin for each OC-SVM _{ik} which is trained with samples of the i -th class $\theta_i, i=1, \dots, n$, provided by the source of information $S_k, k=1, \dots, p$, respectively.

In the multi-class classification framework, OC-SVM classifier is extended. Therefore, the posterior probability $P_k(\theta_i/x)$ of any target class $\theta_i, i=1, \dots, n$, of the frame $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ can be directly obtained according the equation (8). Finally, the maximum likelihood (ML) test is used for decision making as follows:

$$x \in \theta_i \text{ if } P_k(\theta_i/x) = \max\{P_k(\theta_j/x), 1 \leq j \leq n\} \quad (10)$$

where x is the pattern test characterized by the source of information $S_k, k=1, \dots, p$.

B. Estimation of Masses

In this paper, the mass functions $gbba$ of evidence $m_k(.)$ issued from the k -th source of information $S_k, k=1, \dots, p$, are estimated using an inspired version of Appriou's model, which is initially defined for two classes [31]. Therefore, the modified version of Appriou's model in DSmT framework over $G^{(i)} = \{\emptyset, \theta_i, \bar{\theta}_i, \theta_i \cup \bar{\theta}_i, \theta_i \cap \bar{\theta}_i\}$, $i=1, \dots, n$, is given as:

$$m_k(\emptyset) = 0, \quad (11)$$

$$m_k(\theta_i) = \frac{(1 - \beta_i) P_k(\theta_i/x)}{(1 + \varepsilon)(1 + P_k(\theta_i/x))}, \quad (12)$$

$$m_k(\bar{\theta}_i) = \frac{1 - \beta_i}{(1 + \varepsilon)(1 + P_k(\theta_i/x))}, \quad (13)$$

$$m_k(\theta_i \cup \bar{\theta}_i) = \frac{\varepsilon}{1 + \varepsilon}, \quad (14)$$

$$m_k(\theta_i \cap \bar{\theta}_i) = \frac{\beta_i}{1 + \varepsilon}, \quad (15)$$

where $\varepsilon \geq 0$ is a tuning parameter, and β_i is the sum of false accepted rates (FAR) made by the OC-SVM $_{ik}, k = 1, \dots, p$, classifiers, which are trained with p sources of information, respectively. Furthermore, $\beta_i/(1 + \varepsilon)$ is used to quantify the belief for conflicting region, and $\varepsilon/(1 + \varepsilon)$ is used to quantify the belief that the pattern x belong to the subset $\Theta^{(i)}, i = 1, \dots, n$. Therefore, the value of ε is fixed here to 0.001.

C. Combination of Masses

In order to manage the conflict generated from p sources of information (i.e. $S_k, k = 1, \dots, p$), the global combined masses are computed over the subsets $G^{(i)}, i = 1, \dots, n$, as follows:

$$\begin{cases} m_c(A) \equiv \frac{1}{n} \times m_c^{(i)}(A) \\ \text{with } A \in G^{(i)}, i = 1, \dots, n, \end{cases} \quad (16)$$

where $1/n$ represents normalization factor that is introduced in DSMT framework in order to respect the normality condition of masses over the set of all focal elements $F = \{G^{(i)}, i = 1, \dots, n\}$ (i.e. $\sum_{A \in F} m_c(A) = 1$), $m_{PCR6}^{(i)}(\cdot)$ is

the partial combined mass using the PCR6 combination rule, such that $m_{PCR6}^{(i)}(A) = [m_1 \oplus \dots \oplus m_p](A)$, and \oplus represents the combination operator, which is composed of both conjunctive and redistribution terms of the PCR6 rule, when dealing with DSMT framework.

D. Decision Rule

Combination of evidences using the proposed combination scheme yields the combined belief and a decision making is made using the statistical classification technique. First, the combined beliefs are converted into probability measure using a probabilistic transformation, called DSMP, that maps a belief measure to a subjective probability measure [32] defined as:

$$DSMP_\varepsilon(A_i) = \begin{cases} 0 & \text{if } A_i \in \emptyset, \\ \sum_{A_j \in F} \frac{W_j}{T_j} m_c(A_j) & \text{otherwise,} \end{cases} \quad (17)$$

where

$$W_j = \sum_{\substack{A_k \subseteq A_i \cap A_j \\ C_M(A_k)=1}} m_c(A_k) + \varepsilon C_M(A_i \cap A_j), \quad (18)$$

$$T_j = \sum_{\substack{A_k \subseteq A_j \\ C_M(A_k)=1}} m_c(A_k) + \varepsilon C_M(A_j) \quad (19)$$

$\varepsilon \geq 0$ is a tuning parameter and F corresponds to the set of all focal elements including eventually all the integrity constraints (if any, i.e. $\exists i \in \{1, \dots, n\}, \theta_i \cap \bar{\theta}_i \equiv \emptyset$) of the model M ($F = \{G^{(i)}, \forall i \in \{1, \dots, n\}, \theta_i \cap \bar{\theta}_i \equiv \emptyset\}$ for Shafer's model and $F = \{G^{(i)}, \forall i \in \{1, \dots, n\}, \alpha_i \equiv \theta_i \cap \bar{\theta}_i \neq \emptyset\}$ for all the paradoxical hypotheses); $C_M(A_k)$ denotes the DSm cardinal of the set A_k [14].

In the context of some particular multi-class classification problems, the simple classes θ_i are truly exclusive and Shafer's model is adopted. Therefore, the $DSMP_\varepsilon(\theta_i)$ probability of any element $\theta_i, i = 1, \dots, n$, of the frame $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ can be directly obtained according the following equation:

$$DSMP_\varepsilon(\theta_i) = \frac{m_c(\theta_i) + (m_c(\theta_i) + \varepsilon) \sum_{\substack{A_j \in F \\ A_j \supset \theta_i \\ C_M(A_j) \geq 2}} \frac{m_c(A_j)}{\sum_{\substack{A_k \in F \\ A_k \subset A_j \\ C_M(A_k)=1}} m_c(A_k) + \varepsilon C_M(A_j)}}{m_c(\theta_i) + (m_c(\theta_i) + \varepsilon) \sum_{\substack{A_j \in F \\ A_j \supset \theta_i \\ C_M(A_j) \geq 2}} \frac{m_c(A_j)}{\sum_{\substack{A_k \in F \\ A_k \subset A_j \\ C_M(A_k)=1}} m_c(A_k) + \varepsilon C_M(A_j)}}. \quad (20)$$

In this manner, the combined belief assignment is transformed into a probability measure so that the statistical classification approach is applied for computing the final decision. Finally, the DSMP-based maximum likelihood (ML) test is used for decision making as follows:

$$x \in \theta_i \text{ if } DSMP_\varepsilon(\theta_i) = \max\{DSMP_\varepsilon(\theta_j), 0 \leq j \leq n\} \quad (21)$$

where x is the pattern test characterized by p sources of information $S_k, k = 1, \dots, p$, and ε is fixed to 0.001 in the decision measure given by (17).

V. DATABASE AND ALGORITHMS USED FOR VALIDATION

The proposed OC-SVM classifiers are trained using different methods of generating features on a database of the isolated handwritten digits. In this section, we briefly describe the database, the methods used for generating features and the algorithm used for validation of OC-SVM models.

A. Database Description and Performance Criteria

To validate the proposed combination scheme, the well-known US Postal Service (USPS) database is used for handwriting recognition task. This database contains normalized grey-level handwritten digit images of 10 numeral classes, extracted from US postal envelopes. All images are segmented and normalized to a size of 16×16 pixels. There

are 7291 training data and 2007 test data where some of them are corrupted and difficult to classify correctly. For evaluating the performances of the combination scheme, a popular rates are considered, which are the *Recognition Rate (RR)* for each class and *Mean Recognition Rate (MRR)* for all classes.

B. Methods Used for Generating Features

The objective of the features generation step is to underline the relevant information that initially exists in the raw data. Thus, an appropriate choice of the descriptor improves significantly the accuracy of the combination scheme. In this study, we use a collection of popular feature generation methods, which can be categorized into background features, foreground features, geometric features [3].

C. Algorithm Used for Validation of OC-SVM Models

The OC-SVM model is produced for each class according the used descriptor. Hence, the training dataset is partitioned into ten subsets of samples: each one is used as a learning subset to learn the corresponding OC-SVM classifier that operates independently of other. Let N different values of the RBF parameter γ_i , sorted in increasing order, such that $i = 0, 1, \dots, N-1$. The hyper parameters of each OC-SVM model are tuned during the validation phase using the corresponding training subset of samples. In this work, we have allowed until 10 % of error on the training dataset (i.e. percentage of training data considered as outliers $\nu \leq 0.1$) for each target class. Hence, the j -th OC-SVM model is generated under constraint $ER_j < \nu \times 100$, such that ER_j is the corresponding error rate computed during the validation phase. Indeed, the selection of the optimal value $\gamma_{opt}^{(i)}$, of the RBF parameter during the validation phase is performed for each target class $\theta_i, i = 0, \dots, 9$, on the set of all models fulfilling the last condition, using the maximum of the number of support vectors S_v criteria. Consequently, higher the number of support vectors is, the better the information is representative for each class.

VI. EXPERIMENTAL RESULTS

The effectiveness of the proposed combination scheme is demonstrated experimentally by evaluating the recognition performance on all isolated handwritten digits from the test dataset. We perform experiments to select a subset of global complementary sources of information using the proposed extension of OC-SVM into multi-class (MC) classification framework, namely MC-OC-SVM, and then the proposed combination scheme is evaluated in belief function theories framework.

A. Performance Evaluation of the Proposed Descriptors

In these experiments, we compute during the test phase the recognition rate MRR of the MC-OC-SVM classifier using Geometric Features (GF), Foreground Features (FF), Background Features (BF), and the descriptors which result from a concatenation between at least two simple descriptors such as (BF,FF), (BF,GF), (FF,GF), and the (BF,FF,GF)

descriptor. Indeed, the experiment has shown that the appropriate choice of both descriptors and concatenation in order to represent each digit class in the feature generation step provides an interesting recognition performance. In Table II, the MRR vary from one descriptor to another, and the MRR of concatenated descriptors are relatively high compared to those of simple descriptors.

As we can see, it is difficult to improve the recognition performance by a concatenation of features since most of the time the combined descriptors does not take into account the complementary nature of features, which can be exist between both descriptors. Hence, we choose among all descriptors available (i.e. see Table II) only those for which the corresponding MC-OC-SVM classifiers could attain an improvement in the recognition performance. Indeed, BF, FF and GF-based descriptors yield respectively in the experiments (c), (b) and (a) a MRR of 89.50%, 83.75% and 78.90%. When using (FF,GF)-based descriptor in the experiment (f), we obtain a significant improvement in the recognition performance of the MC-OC-SVM classifier from 83.75% until 87.59%. Further, an important gain of 4.31% in the recognition performance, where $MRR = 91.71\%$, is obtained in the experiment (g) when the BF-based descriptor is concatenated to the (FF,GF)-based descriptor to get a new (BF,FF,GF)-based descriptor.

Hence, in the following section we use the three descriptors BF, (FF,GF) and (BF,FF,GF) as global sources of information to feed the OC-SVM classifier of each target class. This allows us to evaluate the recognition performance of the proposed combination scheme and to better exploit the complementary nature, which is obtained from these descriptors. In this way, it is possible to improve the recognition performance when the concatenation of descriptors can fail to provide the correct solution for some specific handwritten digit recognition problems.

B. Performance Evaluation of the Proposed Combination Scheme

In these experiments, we evaluate the recognition performance of the proposed combination scheme using sum, DS and PCR6 rules. In fact, this combination scheme allows to exploit the complementary nature issued from the three sources of information $S_1 \equiv BF, S_2 \equiv (FF, GF), S_3 \equiv (BF, FF, GF)$, and manage the conflict provided from the outputs of OC-SVM classifiers of each target class.

The proposed combination scheme, which uses ten combinations (i.e. a combination per target class), consists to measure ten values of conflict $K_{ci}, i = 1, \dots, 9$, for each isolated handwritten digit from the test dataset. In the context of recognition of isolated handwritten digits, the conflicting regions of each handwritten digit test are modeled by the ten paradoxical classes $\theta_i \cap \bar{\theta}_i, i = 0, \dots, 9$. Consequently, we introduce an integrity constraint $\theta_i \cap \bar{\theta}_i, i = 0, \dots, 9$, for each corresponding combination. In Table III, each line presents the FAR error computed for each class of handwritten digits using the three sources of information S_1, S_2 and S_3 . Indeed, the

TABLE II
RECOGNITION RATES OF THE MC-OC-SVM CLASSIFIER USING DIFFERENT METHODS OF GENERATING FEATURES

Descriptor	Simple Descriptor				Concatenated Descriptor		
	(a) GF	(b) FF	(c) BF	(d) (BF,FF)	(e) (BF,GF)	(f) (FF,GF)	(g) (BF,FF,GF)
MRR (%)	78.90	83.75	89.50	90.70	91.90	87.59	91.71

proposed OC-SVM models are optimized and trained using the training dataset to compute, during the validation phase, the error parameter $\beta_i, i = 0, \dots, 9$ shown in the last column of this table, as the sum of FAR errors which is subsequently used for each combination. As we can see here, the 8-th OC-SVM classifier provides the higher value of error of FAR, where $\beta_8 = 6.69\%$, for each source of information.

TABLE III
FAR ERRORS PROVIDED BY THE TREE SOURCES OF INFORMATION FOR EACH CLASS

Class	FAR (%)			β_i (%)
	Source 1	Source 2	Source 3	
0	0.07	0.00	0.00	0.07
1	0.10	0.06	0.06	0.22
2	0.67	1.14	0.18	1.99
3	0.24	0.24	0.11	0.59
4	0.02	0.00	0.00	0.02
5	0.60	0.64	0.21	1.45
6	0.02	0.33	0.14	0.49
7	0.03	0.00	0.00	0.03
8	2.48	2.88	1.33	6.69
9	0.08	0.14	0.06	0.28

For better comparison, recognition results corresponding to the combination of the three sources S_1, S_2 and S_3 by sum, DS and PCR6 rules are respectively given in Table IV. The proposed combination scheme using the DS rule yields a MRR of 92.76% corresponding to an improvement of 1.05%. While sum rule decreases the MRR to 92.68%. This is due to the direct estimation technique of masses which assigns the confidences only to the simple classes in probabilistic theory framework. Hence, the sum rule couldn't handle managing correctly the conflict generated from the three sources. Furthermore, the experimental results show that the DS rule is not able to handle most of the conflicting cases between the three sources. Hence, the DST is not appropriate to solve our problem of handwritten digit recognition into the multi-class classification framework. Indeed, the use of DS rule in the combination scheme allows redistributing the beliefs through a simple normalization by $(1 - K_{ci}, i = 0, 1, \dots, 9)$ in the combination process of masses. However, when responses of OC-SVM classifiers are less complementary, both sum and DS rules do not provide reliable decision. In Table IV, an improvement of RR when using PCR6 rule compared to both sum and DS rules were obtained for all classes of handwritten digits except those of θ_1, θ_2 and θ_8 . This is because there are some digits belonging to these classes which are wrongly characterized by the combined sources S_1, S_2 and S_3 . In other words, the PCR6 combination based rule is not reliable when the complementary information provided from the sources of information is wrongly preserved. In order to get a higher MRR, the combined sources of information should provide complementary information. The proposed combination scheme with PCR6 rule yields the best MRR of 95.03% when

TABLE IV
COMPARISON OF THE RECOGNITION PERFORMANCE OF SUM, DS AND PCR6 COMBINATION RULES USING THREE SOURCES (S_1, S_2, S_3)

Class	Sources			Combination Rules		
	S_1	S_2	S_3	Sum	DS	PCR6
0	81.05	80.22	86.07	86.35	86.35	94.15
1	85.55	94.67	96.57	95.06	95.05	96.20
2	92.78	92.78	96.39	95.36	95.87	94.84
3	81.09	75.00	85.97	89.02	89.63	92.07
4	94.44	87.87	91.91	93.94	93.43	95.95
5	91.81	84.37	89.37	91.87	91.87	95.00
6	92.30	91.12	94.67	94.08	93.49	97.63
7	91.09	91.78	91.78	93.15	93.84	93.83
8	93.29	94.51	95.73	98.17	98.17	95.12
9	91.52	83.61	88.70	89.83	89.83	95.48
MRR (%)	89.50	87.59	91.71	92.68	92.76	95.03

combining the three sources S_1, S_2 and S_3 all together. Indeed, the PCR6 rule allows an efficient redistribution of the partial conflicting mass only to the elements involved in the partial conflict (i.e. the target class θ_i and their complementary class $\bar{\theta}_i, i = 0, 1, \dots, 9$). After redistribution, the combined mass is transformed into the DS_m probability and the DS_mP-based ML test is used for decision making. Finally, the proposed combination scheme using PCR6 rule in DS_mT framework is the most stable across all experiments whereas recognition rates pertaining to DS combination rule vary significantly.

VII. CONCLUSION AND FUTURE WORK

In this paper, an effective combination scheme of one-class classifiers in a general belief function framework has been proposed. The OC-SVM classifiers can be incorporated as an intelligent learning technique for reducing the number of focal elements. This scheme consists in using of a subset of global complementary sources of information to feed the OC-SVM classifiers corresponding to each target class, which allows decomposing a n -class problem into a series of n -combination, while providing n -calibrated outputs into the multi-class framework. Therefore, this allows us to reduce drastically the computational complexity of the combination process and to extend specially the applicability of DS_mT into the multi-class classification framework. Experimental results show that the proposed combination scheme with PCR6 rule yields the best performance on the handwritten digit recognition application compared to the sum rule and DS rule even when the individual MC-OC-SVM multi-classifications provide uncalibrated outputs.

In continuation to the present work, the next objectives consist to adapt the use of the evidence supporting measure of similarity (ESMS) criteria to select complementary sources of information for each target class using the same proposed combination scheme in order to attempt to improve the RR and MRR.

ACKNOWLEDGMENT

The authors would like to thank Dr. A. Martin, Dr. F. Smarandache and Dr. J. Dezert for their valuable suggestions.

REFERENCES

- [1] R. O. Duda, P. E. Hart, and D. G. Stork. Pattern Classification. 2nd edition, Wiley Interscience, New York, 2001.
- [2] M. Cheriet, N. Khama, C.-L. Liu, and C. Y. Suen. Character Recognition Systems: A Guide for Students and Practitioner. John Wiley & Sons, 2007.
- [3] N. Abbas, Y. Chibani and H. Nemmour, "Handwritten Digit Recognition Based On a DSMT-SVM Parallel Combination," in Proc. 13th International Conference on Frontiers in Handwriting Recognition (ICFHR), Bari, Italy, September 18-20, 2012. <http://www.icfhr2012.uniba.it/paper038.pdf>
- [4] L. A. Zadeh. Fuzzy algorithm. Inform. Contr, 12, 1968, pp. 94-102.
- [5] G. Shafer. A Mathematical Theory of Evidence. Princeton University Press, 1976.
- [6] D. Dubois and H. Prade, "Representation and combination of uncertainty with belief functions and possibility measures," Computational Intelligence, vol. 4, pp. 244-264, 1988.
- [7] F. Smarandache and J. Dezert, Advances and Applications of DSMT for Information Fusion. Rehoboth, NM: Amer. Res. Press, 2004.
- [8] F. Smarandache and J. Dezert, Advances and Applications of DSMT for Information Fusion. Rehoboth, NM: Amer. Res. Press, 2006.
- [9] F. Smarandache and J. Dezert, Advances and Applications of DSMT for Information Fusion. Rehoboth, NM: Amer. Res. Press, 2009.
- [10] M. Vatsa, R. Singh, A. A. Ross and Noore, "On the Dynamic Selection in Biometric Fusion Algorithms," IEEE Transaction on Information Forensics and Security, vol. 5, no. 3, pp. 470-479, 2010.
- [11] N. Abbas and Y. Chibani, "SVM-DSMT combination for the simultaneous verification of off-line and on-line handwritten signatures," International Journal of Computational Intelligence and Applications (IJCIA), vol. 11, no. 3, 2012.
- [12] M. Tombak, A. Isotamm and T. Tamm, "On logical method for counting Dedekind numbers," Lect. Notes on Comp.Sci., 2138, pp. 424-427, Springer-Verlag, 2001. http://www.cs.ut.ee/people/m_tombak/publ.html.
- [13] J. Dezert and F. Smarandache, "The generation of the hyper-power sets," Chap. 2, pp. 37-48, in Advances and Application of DSMT for Information Fusion. Rehoboth, NM: Amer. Res. Press, 2004.
- [14] J. Dezert and F. Smarandache, "Partial ordering on hyper-power sets," Chap. 3, pp. 49-60, in Advances and Application of DSMT for Information Fusion. Rehoboth, NM: Amer. Res. Press, 2004.
- [15] R. Kennes, "Computational Aspect of the Möbius Transformation of Graphs," IEEE Transaction on Systems, Man, and Cybernetics – Part A: Systems and Humans, vol. 22, no. 2, pp. 201-223, 1992.
- [16] Ph. Smets, "The application of matrix calculus for belief functions," International Journal of Approximate Reasoning, vol. 31, pp. 1-30, 2002.
- [17] T. Denoeux, "Inner and outer approximation of belief structures using a hierarchical clustering approach," Int. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 9, no. 4, pp. 437-460, 2001.
- [18] P. Djiknavorian and D. Grenier, "Reducing DSMT hybrid rule complexity through optimization of the calculation algorithm," Chap. 15, pp. 345-440, in Advances and Application of DSMT for Information Fusion. Rehoboth, NM: Amer. Res. Press, 2006.
- [19] A. Martin and C. Osswald, "A new generalization of the proportional conflict redistribution rule stable in terms of decision," Chap. 2, pp. 69-88, in Advances and Application of DSMT for Information Fusion. Rehoboth, NM: Amer. Res. Press, 2006.
- [20] A. Martin, "Implementing general belief function framework with a practical codification for low complexity," Chap. 7, pp. 217-273, in Advances and Application of DSMT for Information Fusion. Rehoboth, NM: Amer. Res. Press, 2009.
- [21] J. Dezert and F. Smarandache, "Combination of beliefs on hybrid DSMT models," Chap. 4, pp. 61-103, in Advances and Application of DSMT for Information Fusion. Rehoboth, NM: Amer. Res. Press, 2004.
- [22] X. Li, J. Dezert, F. Smarandache, X. Huang, "Evidence Supporting Measure of Similarity for Reducing the Complexity in Information Fusion," Information sciences, vol. 181, no. 10, pp. 1818-1835, 2011.
- [23] F. Smarandache and J. Dezert, "Proportional Conflict Redistribution Rules for Information Fusion," Chap. 1, pp. 3-68, in Advances and Application of DSMT for Information Fusion. Rehoboth, NM: Amer. Res. Press, 2006.
- [24] J. Dezert, "Fondations for a new theory of plausible and paradoxical reasoning," Information & Security Journal, vol. 9, pp. 13-57, 2002.
- [25] Y. Guermeur, A. Elisseeff and H. Paugam-Moisy, "Estimating the sample complexity of a multi-class discriminant model," in Proc. Industrial Conference on Artificial Neural Networks, pp. 310-315, 1999.
- [26] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola and R. C. Williamson, "Estimating the support of a high dimensional distribution," Neural Computation, vol. 13, no. 7, pp. 1443-1472, 2001.
- [27] V. N. Vapnik, The Nature of Statistical Learning Theory. Springer. New York, 1995.
- [28] C. M. Bergamini, L. S. Oliveira, A. L. Koerich, and R. Sabourin, "Combining different biometric traits with one-class classification," Signal Processing, vol. 89, no. 11, pp. 2117-2127, 2009.
- [29] Y. Guerbai, Y. Chibani, and N. Abbas "One-Class versus Bi-Class SVM Classifier for Off-line Signature Verification," in The 3rd Intl. Conf. on Multimedia Computing and Systems, Tangier, Morocco, May 2012.
- [30] Q.-A. Tran, X. Li and H. Duan, "Efficient performance estimate for one-class support vector machine," Pattern Recognition Letters, vol. 26, pp. 1174-1182, 2005.
- [31] A. Appriou, "Probabilités et incertitude en fusion de données multisenseurs," Revue Scientifique et Technique de la Défense, vol. 11, pp. 27-40, 1991.
- [32] J. Dezert and F. Smarandache, "Transformations of belief masses into subjective probabilities," Chap. 3, pp. 85-136, in Advances and Application of DSMT for Information Fusion. Rehoboth, NM: Amer. Res. Press, 2009.